DOCUMENT RESUME

ED 080 085 HE 004 487

AUTHOR	Oliver, Robert M.
TITLE	Models for Predicting Gross Enrollments at the University of California.
INSTITUTION	California Univ., Berkeley. Ford Foundation Program for Research in Univ. Administration.
SPONS AGENCY	Ford Foundation, New York, N.Y.
REPORT NO	RR-68-3
PUB DATE	Aug 68
NOTE	29p.
AVAILABLE FROM	Ford Foundation, 2288 Fulton Street, Berkeley, California 94720
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*College Students; *Educational Planning; Educational Research; *Enrollment Trends; *Higher Education; Models; Research; Research Projects
IDENTIFIERS	*University of California

ABSTRACT

The purpose of this report is to discuss and compare two mathematical models for predicting student enrollments at the University of California. One has been proposed in the scientific literature and the second has been used by the state of California since 1963 to forecast student enrollments. The specific problems addressed in this report are the prediction of gross enrollments, i.e., freshmen, sophomores, etc., for a particular campus of the University as a whole. Although the experimental data is restricted to undergraduates, the discussion and conclusions are probably appropriate to graduate levels as well. (Author)

ED 080085

FORD FOUNDATION PROGRAM FOR RESEARCH IN UNIVERSITY ADMINISTRATION

Office of the Vice President—Planning
University of California

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

HE 004487

FILMED FROM BEST AVAILABLE COPY

FORD GRANT NO. 680-0267A

RESEARCH DIRECTORATE

Charles J. Hitch	President, University of California
Frederick E. Balderston	Professor of Business Administration Chairman, Center for Research in Management Science University of California, Berkeley Academic Assistant to the President
George B. Weathersby	Associate Director, Office of Analytical Studies University of California

OFFICE ADDRESS

2288 Fulton Street
Berkeley, California 94720
(415) 642-5490

(List of Available Publications on Inside Back Cover)

ED 080085

MODELS FOR PREDICTING GROSS ENROLLMENTS AT THE UNIVERSITY OF CALIFORNIA

Robert M. Oliver

Research Report No. 68-3
August 1968

TABLE OF CONTENTS

I. INTRODUCTION

1. Background 1

II. ENROLLMENT FORECASTS

1. Two Mathematical Models 4

2. One-Period Forecasts 8

3. Admissions, Enrollments and Long Term Forecasts 13

4. The Grade Progression Ratio Method 17

III. NUMERICAL EXAMPLES AND COMPARISON OF FORECASTS

1. Berkeley Campus Forecasts, 1960-1966

BIBLIOGRAPHY

MODELS FOR PREDICTING GROSS ENROLLMENTS AT THE UNIVERSITY OF CALIFORNIA

I. INTRODUCTION

1. Background

The purpose of this report is to discuss and compare two mathematical models for predicting student enrollments at the University of California. One has already been proposed in the scientific literature and we will refer to it as the Gani-Young-Almond model (GYA), while the second has been used by the State of California since 1963 to forecast student enrollments; we will refer to the latter as the Grade Progression Ratio method (GPR).

Enrollment forecasts are required for different purposes. At the departmental level, for example, they are used to predict faculty workloads and make faculty, classroom and advising assignments. At the statewide level forecasts are used for overall budget and planning purposes. In this report we are primarily concerned with forecasts of the latter type in the absence of quota restrictions on total enrollments. The effect of such quotas will be studied at a later time.

Forecasts differ not only in the degree of fineness to which they predict various categories of students, but they often refer to different periods of time. Predictions of gross enrollments may refer to an upper division enrollment during an academic year, while the forecasts of departmental majors may only be useful if they refer to class enrollments at the beginning of each quarter. Besides the usual statistical question

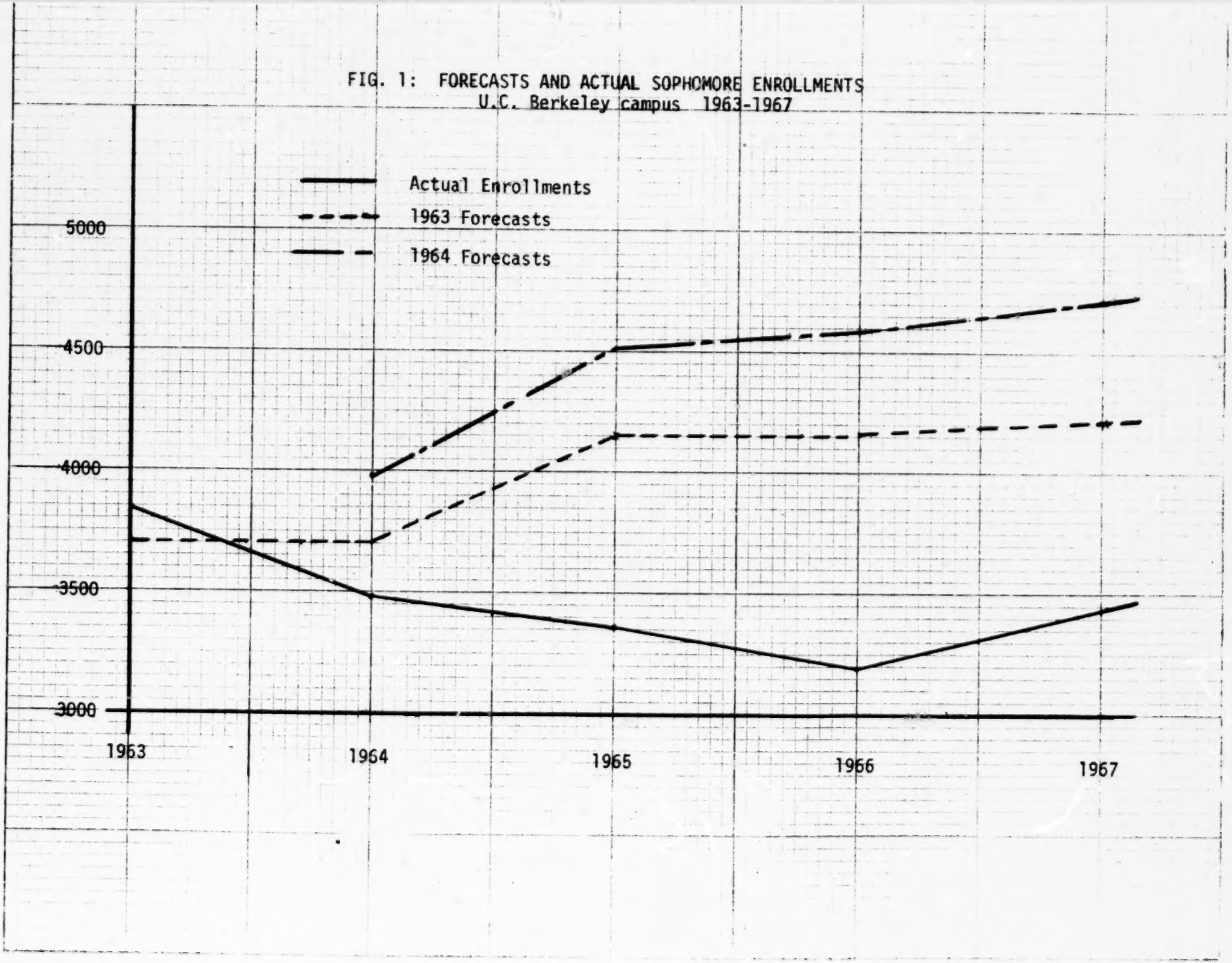
of the reliability of any forecast, an item of major importance is the degree to which numbers obtained by aggregating grades and time periods in one model is consistent with numbers obtained from a separate model that predicts gross enrollment figures in an unaggregated form.

A second item of importance is the clear distinction between the variables being predicted and those identified as policy variables. As we will point out in later sections of this report, this distinction is particularly important in periods where enrollment quotas are imposed and one attempts to find admission, redirection and student reclassification policies that maintain these quotas or other restrictions imposed by a university administration.

The specific problems that we address in this report are the prediction of gross enrollments, i.e., freshmen, sophomores, etc., for a particular campus or the University as a whole. Although we restrict our experimental data to undergraduates, the discussion and conclusions are probably appropriate to graduate levels as well.

Figure 1 is a plot of the actual numbers and forecasts of Berkeley campus enrollments in the period 1963-1967. The solid line refers to actual enrollments, the dashed lines to forecasts made before the beginning of Fall 1963 and Fall 1964 by the Department of Finance of the State of California.

FIG. 1: FORECASTS AND ACTUAL SOPHOMORE ENROLLMENTS
U.C. Berkeley campus 1963-1967



II. ENROLLMENT FORECASTS

1. Two Mathematical Models

Denote by $X_i(t)$ the number of enrolled students in grade i at the beginning of time period t. The subscript index might, for example, refer to sophomores, and the parameter t might refer to the beginning of the fall quarter 1967. Denote by $Y_i(t)$ the new admissions to grade i during the t^{th} time period. It is the purpose of enrollment forecast models to make a prediction or estimate of $X_i(t+1)$ on the basis of certain historical information and past trends on enrollment and admission data.

One model that has evoked considerable interest assumes that the fraction which leave grade i and go to grade j is a fraction p_{ij} such that

$$X_j(t+1) = \sum_{i=1}^{m} X_i(t)\, p_{ij} + Y_j(t+1) \qquad (1)$$

where it is understood that for some i, $\sum_{j=1}^{m} p_{ij} < 1$. The p_{ij}'s may themselves be time dependent.

Gani (1963) used such a model for predicting gross enrollments in the Australian university system. His statistical data seems to indicate that it is reasonable to assume that transitions between grades have a fixed probability over time for time periods of the order of five years. In 1965 Gani adopted a revised model for use at Michigan State University which also took into account (i) the number

of credits completed by the student, and (ii) the possibility of transfers between majors.

The GYA model has the important feature that contributions to the enrollments in one grade are identified by their origins (prior grades, return to the same grade, new admissions, etc.) and are added to give a total enrollment figure. Secondly, it has the appealing feature that the fractions p_{ij} can be interpreted as transition probabilities and thus allows one to adopt useful results from the theory of Markov Chains even though the process itself may not be Markovian. Thirdly, the conditional short-term forecasts given today's enrollments have an intuitively correct structure, and finally, the method has some experimental evidence to support it.

A second model that has been used by the State of California Department of Finance for predicting gross campus enrollments at the University of California is called the Grade Progression Ratio method. Although there are no published reports to document its mathematical structure, the method is based on defining progression ratios a_{11} and $a_{j,j+1}$ for the j^{th} grade and then using these ratios to predict enrollments $Z_j(t)$ for future periods by means of the system of first-order difference equations,

$$\begin{aligned} Z_1(t+1) &= a_{11}\, Z_1(t) + Y_1(t) \\ Z_{j+1}(t+1) &= a_{j,j+1}\, Z_j(t) + Y_{j+1}(t+1) \end{aligned} \qquad (2)$$

$a_{11} = p_{11}$ is the fraction of freshmen who return to that grade in the next period.

The ratio $a_{j,j+1}$ has the interpretation that it is the fraction of continuing students in grade j+1 (total minus new admissions) relative to total enrollments of a lower grade at the beginning of a previous period. In many of the applications that we have seen these coefficients also vary with time.

Throughout the remainder of this report we will make the further assumption that student flows between grades are only of three types: (i) to the same grade, (ii) to the next higher grade, or (iii) by departure from the university system. In this case Equation (1) can be specialized to yield the result

$$X_{j+1}(t+1) = X_j(t)\, p_{j,j+1} + X_{j+1}(t)\, p_{j+1,j+1} + Y_{j+1}(t+1) \qquad (3)$$

In other words, enrollees in grade (j+1) at the beginning of time period (t+1) come from grade j or grade j+1 in the previous period or represent new admissions. In the absence of quota restrictions on new admissions or total class size, Equation (3) is assumed to represent the underlying stochastic process and the problem is reduced to one of

(a) estimating $p_{j,j}$ and $p_{j,j+1}$ terms

(b) estimating $Y_j(t)$ terms

(c) recursively computing forecasts and estimates of fluctuations from Equation (3)

(d) establishing time periods which are natural to the forecasting process

and

(e) making experimental comparisons of forecast and real data.

2. One-Period Forecasts

In this section we consider the distribution of returning and new students given today's enrollments. Given the value $X_j(t)$ for the enrollments in grade j at time t, the number that return to the same grade, that advance to the next higher grade or that drop out in the succeeding time period are multinomially distributed. To illustrate, we denote by $X_{j,j}$ the random number that do not advance, by $X_{j,j+1}$ the number that advance from grade j to j+1 in one time period and by $X_{j,0}$ the number that drop out. The probability that there are x returning to grade j, y advancing and z leaving is

$$\Pr\{X_{j,j} = x,\ X_{j,j+1} = y,\ X_{j,0} = z \mid X_j = n\} = \frac{n!}{x!y!z!} (p_j)^x (q_j)^y (r_j)^z \tag{4}$$

$$p_j + q_j + r_j = 1$$
$$x + y + z = n$$

where, to simplify notation, we have deleted the time parameter and substituted p_j for $p_{j,j}$, q_j for $p_{j,j+1}$ and r_j for the drop-out probability, $1 - p_{j,j} - p_{j,j+1}$.

The probability distribution for the total number of continuing students, i.e., those remaining in grade plus those advanced from a lower grade, given the actual enrollments in the previous period is

$$\Pr\{X_{j,j+1} + X_{j+1,j+1} = k \mid X_j = m,\ X_{j+1} = n\}$$

$$= \sum_{i=0}^{k} \frac{m!}{i!(m-i)!} \frac{n!}{(k-i)!(n-k-i)!} \tag{5}$$

$$\times \left[q_j^{\,i}(1-q_j)^{m-i}\,(p_{j+1})^{k-i}\,(1-p_{j+1})^{n-k-i}\right]$$

It follows from (5) or direct expectation arguments that the conditional one-period expectations are

$$E[X_{j+1} \mid m, n] = mq_j + np_{j+1} + E[Y_{j+1}] \tag{6}$$

where, once again, the time parameter is suppressed. With the further assumption that Y_j's are statistically independent of previous X_j's, the conditional variance of enrollments in the next period is

$$\mathrm{Var}\,[X_{j+1} \mid m, n] = mq_j(1-q_j) + np_{j+1}(1-p_{j+1}) + \mathrm{Var}\,[Y_{j+1}] \tag{7}$$

given m students in grade j and n students in grade j+1.

The variance to mean ratio for one-period forecasts of continuing students given that there are m students in grade j, n in grade j+1 is

$$1 - \mathrm{Max}\,(q_j, p_{j+1}) < \frac{mq_j(1-q_j) + np_{j+1}(1-p_{j+1})}{mq_j + np_{j+1}} \tag{8}$$

$$< 1 - \mathrm{Min}\,(q_j, p_{j+1})$$

with the GYA model. This variance to mean ratio may be reduced or increased by new admissions and transfer of students as is indicated

by the ratio,

$$\frac{mq_j(1 - q_j) + np_{j+1}(1 - p_{j+1}) + \text{Var}\,[Y_{j+1}]}{mq_j + np_{j+1} + E[Y_{j+1}]}, \qquad (9)$$

for next period enrollments. When new admissions can be forecast exactly, $\text{Var}\,[Y_{j+1}] = 0$; the denominator of (9) is larger than (8) and the variance to mean ratio of enrollments is considerably smaller than the corresponding figure for continuing students. If the $\{Y_j\}$ process is characterized by a sequence of independent Poisson variables, then the ratio of (9) is larger than the ratio in (8).

Consider, for example, the case of sophomore enrollments in the year 1961 at the Berkeley campus. When we assume that $q_1 = p_{12} = 0.58$, $p_2 = p_{22} = 0.23$, $m = 3843$, $n = 3445$, the variance to mean ratio for continuing students in 0.51. The number of new sophomore admissions in 1961 is 751 students. If this number were known with certainty, then the variance to mean ratio of the one-period enrollment forecast is reduced from 0.51 to 0.41. If, on the other hand, the new sophomore admissions for fall 1960 are Poisson with $\text{Var}\,[Y_2] = E[Y_2] = 751$, then the variance to mean ratio for the one-period forecast increases to the value of 0.61. The usual method for reporting forecasts is to give a figure for the mean plus or minus two standard deviations. In the first case we would predict a figure of 3772 ± 79 students enrolled in the sophomore class of 1961 while in the latter case we would obtain 3772 ± 96 students.

Since we are unable to find published accounts for the GPR model, it is difficult to guess what the underlying stochastic process might be for continuing students. It may be reasonable to assume that each undergraduate at grade $j \geq 2$ leads to $k = 0, 1, 2, 3, \ldots$ continuing undergraduates at the next higher grade with probability distribution p_k having mean and variance

$$\mu_j = \sum_k k p_k; \qquad \sigma_j^2 = \sum_k (k - \mu)^2 p_k . \tag{10}$$

Our data seems to suggest that for the transitions from grade j to grade j+1

$$\mu_j = a_{j,j+1}$$

Using conditional expectation arguments and the assumption that each student acts independently of all others, we obtain

$$E[Y_{j+1}(t+1) \mid Y_j(t) = m] = m a_{j,j+1} \tag{11}$$

and

$$\text{Var}\,[Y_{j+1}(t+1) \mid Y_j(t) = m] = \sum_1^m \sigma_j^2 = m\sigma_j^2 . \tag{12}$$

It is important to notice that with such a model the one-period expectations are unaffected by the number of enrollments, say n, in grade j+1 at time period t. Furthermore, the variance to mean ratio for continuing students is independent of m as distinct from the results of (8) which

depend on both m and n. These are some of the reasons why one might be tempted to rely more heavily on the GYA than on the GPR model.

3. Admissions, Enrollments and Long Term Forecasts

If Equation (1) represents the underlying stochastic process, it is possible to recursively compute expectations and variances of enrollments in future periods. If we write (1) in matrix form,

$$X(t+1) = P\,X(t) + Y(t+1) \tag{13}$$

we compute recursively on t to obtain

$$X(t+1) = P^{t+1}\,X(0) + \sum_{j=0}^{t} P^{j}\,Y(t-j+1) \tag{14}$$

It is well known that elements of higher powers of P decrease geometrically with t; hence the initial enrollments represented by X(0) in the matrix equation of (14) have vanishingly small effect on distant enrollment forecasts. In the case of a university system and where few students jump grades, the typical lifetime of an undergraduate student is of the order of four or five years; thus the major contribution to X(t+1) is due to new admissions in one, two or three years just prior to the forecast date. Notice, for example, that the second power of P is

$$P^2 = \begin{vmatrix} p_{11}^2 & 0 & 0 & 0 \\ p_{12}(p_{11}+p_{22}) & p_{22}^2 & 0 & 0 \\ p_{12}p_{23} & p_{23}(p_{22}+p_{33}) & p_{33}^2 & 0 \\ 0 & p_{23}p_{34} & p_{34}(p_{33}+p_{44}) & p_{44}^2 \end{vmatrix}$$

Since the magnitude of the typical grade advance probability at the University of California lies between 1/2 and 3/5 while the return to grade probability is less than or equal to 1/10, terms on the diagonal of powers of P tend to become small rapidly in comparison with non-zero terms below the diagonal and the advance of students through grades is rapid in comparison to a typical management hierarchy where the diagonal terms tend to be much larger.

The calculation of expected enrollments at the beginning of period t is straightforward: we simply substitute E[X(t)] for X(t) and E[Y(t)] for Y(t) in (13) or (14). To get some idea of the fluctuations that we can expect, it is useful to compute the variances of X(t). To do this we make use of the result that the unconditional variance of X(t+1) is related to the conditional variance and expectation of X(t) by means of the formula

$$\text{Var}\,[X(t+1)] = E[\text{Var}\,X(t+1) \mid X(t)] + \text{Var}\,[E[X(t+1) \mid X(t)]] \qquad (15)$$

Considering enrollments in the freshman class we have from (14)

$$\begin{aligned} E[X_1(1)] &= p_{11}\, X_1(0) + E[Y_1(1)] \\ E[X_1(2)] &= p_{11}^2\, X_1(0) + p_{11}\, E[Y_1(1)] + E[Y_1(2)] \qquad (16) \\ E[X_1(3)] &= p_{11}^3\, X_1(0) + p_{11}^2\, E[Y_1(1)] + p_{11}\, E[Y_1(2)] + E[Y_1(3)] \\ &\vdots \end{aligned}$$

where $X_1(0)$ is known and given. The variance of $X_1(1)$ is obtained from

$$\text{Var}\,[X_1(1) \mid X_1(0)] = p_{11}(1 - p_{11})\,X_1(0) + \text{Var}\,[Y_1(1)]$$

$$E\,[X_1(1) \mid X_1(0)] = p_{11}X_1(0) + E\,[Y_1(1)]$$

Since $E\,[X_1(0)] = X_1(0)$ and $\text{Var}\,[X_1(0)] = \text{Var}\,[E\,[Y_1(1)]] = 0$, the unconditional variance of $X_1(1)$ is just equal to our earlier result in Equation (7)

$$\text{Var}\,[X_1(1)] = p_{11}(1 - p_{11})\,X_1(0) + \text{Var}\,[Y_1(1)]\,. \qquad (17a)$$

In the next period additional terms enter because

$$\text{Var}\,[X_1(2) \mid X_1(1)] = p_{11}(1 - p_{11})\,X_1(1) + \text{Var}\,[Y_1(2)]$$

$$E\,[X_1(2) \mid X_1(1)] = p_{11}X_1(1) + E\,[Y_1(2)]$$

and the sum of the expected value of the former and the variance of the latter yields the result

$$\begin{aligned}\text{Var}\,[X_1(2)] = {} & p_{11}^{\,2}(1 - p_{11})\,X_1(0) + p_{11}^{\,3}(1 - p_{11})\,X_1(0) \\ & + p_{11}(1 - p_{11})\,E\,[Y_1(1)] + p_{11}^{\,2}\,\text{Var}\,[Y_1(1)] \\ & + \text{Var}\,[Y_1(2)]\end{aligned} \qquad (17b)$$

In this way one can recursively compute $\text{Var}\,[X_j(t)]$; efficient matrix methods for computing them are described by Pollard (1967) and

Bartholomew (1967). For long term forecasts, the effect of fluctuations due to the initial enrollments becomes small and the dominant terms are due to variances in the new admissions of the immediately preceding years and the variances due to uncertainty of continuing students.

In making long term forecasts one of the two cases that usually interests us corresponds to the assumption that $Y_j(t)$ are known exactly; with such an assumption, fluctuations in enrollments are entirely due to the random nature of attritions and uncertainty in a student's progress once he has enrolled. A second case corresponds to Poisson admissions; in this case enrollment fluctuations are due to the superposition of random admissions with the random behavior of students once they are in the system.

Pollard (1967) has shown that if the number of new admissions are sequences of independent Poisson variables, then the students remaining in grade k at the beginning of the next time period are also Poisson distributed. Since the total enrollments in any grade are the sum over all students who have entered in prior years plus new admissions, the total number of enrollments in each grade are also Poisson distributed. The usefulness of this result lies in the fact that the variance and mean value of the Poisson distribution are equal and that it probably represents a realistic estimate of the magnitude of fluctuations in new admissions.

4. The Grade Progression Ratio Method

To illustrate the mathematical structure of the GPR forecasting model we consider forecasts of the four undergraduate grades as obtained from Equation (2):

$$\begin{vmatrix} Z_1(t+1) \\ Z_2(t+1) \\ Z_3(t+1) \\ Z_4(t+1) \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 & 0 \\ a_{12} & 0 & 0 & 0 \\ 0 & a_{23} & 0 & 0 \\ 0 & 0 & a_{34} & 0 \end{vmatrix} \begin{vmatrix} Z_1(t) \\ Z_2(t) \\ Z_3(t) \\ Z_4(t) \end{vmatrix} + \begin{vmatrix} Y_1(t+1) \\ Y_2(t+1) \\ Y_3(t+1) \\ Y_4(t+1) \end{vmatrix}$$

In matrix notation,

$$Z(t+1) = AZ(t) + Y(t+1) \tag{18}$$

The coefficient $a_{11} = p_{11}$ is the fraction of returning freshmen. All other non-zero entries lie below the main diagonal and may be less than, equal to or greater than one. By iteration one obtains the $(t+1)^{st}$ forecast in terms of the new admissions in prior periods and powers of A

$$Z(t+1) = A^{t+1} Z(0) + \sum_{j=0}^{t} A^j Y(t - j+1). \tag{19}$$

Since $a_{11} = p_{11} < 1$, it is possible to discuss the asymptotic character of powers of A even though values for $a_{j,j+1}$ may be greater than one. In fact, Equation (19) is identical to the development of forecasts of $X_j(t)$ in powers of P except that A is substituted for P and Z for X.

The difference lies in the structure of A, its rank, and the fact that it may not be possible to neglect high powers of A. For example, the second and third powers of A are given by:

$$A^2 = \begin{vmatrix} a_{11}^2 & 0 & 0 & 0 \\ a_{12}a_{11} & 0 & 0 & 0 \\ a_{23}a_{12} & 0 & 0 & 0 \\ 0 & a_{23}a_{24} & 0 & 0 \end{vmatrix}$$

$$A^3 = \begin{vmatrix} a_{11}^3 & 0 & 0 & 0 \\ a_{12}a_{11}^2 & 0 & 0 & 0 \\ a_{23}a_{12}a_{11} & 0 & 0 & 0 \\ a_{23}a_{34}a_{12} & 0 & 0 & 0 \end{vmatrix}$$

and it is likely that the product $a_{23}a_{34}a_{12}$ is large enough so that new freshman admissions of earlier years may dominate all other terms in the forecasts for seniors. For large t the first column entries of A^t are

$$a_{11}^t$$
$$a_{12}a_{11}^{t-1}$$
$$a_{23}a_{12}a_{11}^{t-2}$$
$$a_{23}a_{34}a_{12}a_{11}^{t-3}$$

while all other entries are zero. Hence for large t the typical contribution to the forecast of grade j is due to a fraction of earlier freshman enrollments.

Although we do not make any attempt to discuss the statistical problem of estimating p_{ij} or $a_{j,j+1}$ terms, it has been common practice by the State of California to estimate values for $a_{j,j+1}$ on the basis of recent observations of the random enrollment data, namely, to define a time-dependent ratio

$$a_{j,j+1}(t+1) = \frac{X_{j+1}(t+1) - Y_{j+1}(t+1)}{X_j(t)} \tag{20}$$

where X and Y now refer to actual realizations of the enrollments. Suppose that Equation (3) represents the underlying stochastic process; then the enrollment in grade j+1 at time period t+2 is

$$X_{j+1}(t+2) = X_j(t+1)\, p_{j,j+1} + X_{j+1}(t+1)\, p_{j+1,j+1} + Y_{j+1}(t+2) . \tag{21}$$

If, in making forecasts, we were to use Grade Progression Ratios generated by (3) and (20), then

$$\begin{aligned} a_{j,j+1}(t+1) &= \frac{X_j(t)\, p_{j,j+1} + X_{j+1}(t)\, p_{j+1,j+1}}{X_j(t)} \\ &= p_{j,j+1} + \frac{X_{j+1}(t)}{X_j(t)}\, p_{j+1,j+1} . \end{aligned} \tag{22}$$

Generating the sequence of numbers obtained by substituting (22) into

$$Z_{j+1}(t+2) = Z_j(t+1)\, a_{j,j+1}(t+1) + Y_{j+1}(t+2) , \tag{23}$$

yields the result

$$Z_{j+1}(t+2) = X_j(t+1)\, p_{j,j+1} + \left|\frac{X_j(t+1)}{X_j(t)}\right| X_{j+1}(t)\, p_{j+1,j+1} + Y_{j+1}(t+2) \quad (24)$$

which, though similar to Equation (21), leads to substantially different values when enrollments $X_i(t)$ are time-dependent.

There are two obvious cases where the sequences generated by (24) and (21) do, in fact, agree. If $p_{j+1,j+1}$ terms vanish, then the only contribution to $Z_{j+1}(t+2)$ is due to continuing students from a lower grade and new admissions. In this case the sequence $\{X_j(t)\}$ and $\{Z_j(t)\}$ are identical.

Furthermore, if an equilibrium has been reached in the sense that

$$Y_j(t) \to Y_j; \qquad X_j(t) \to X_j \qquad \text{independently of } t$$

then the ratio in square brackets in (24) is one and we obtain the simpler time-independent recursion

$$Z_{j+1} = X_j\, p_{j,j+1} + X_{j+1}\, p_{j+1,j+1} + Y_{j+1} \quad (25)$$

which agrees with (21) upon deletion of the time parameter t and substitution of Z_{j+1} for X_{j+1}.

III. NUMERICAL EXAMPLES AND COMPARISON OF FORECASTS

1. Berkeley Campus Forecasts, 1960-1966

Table I lists two forecasts and the actual observed enrollments of undergraduate students on the Berkeley campus for the period 1962-1966. The top entry in each cell is computed by the GYA model with the new admission data shown in Table II. It was assumed that the transition probabilities for Fall to Spring and Spring to Fall semesters were those given by the P_1 (Fall to Spring) and P_2 (Spring to Fall) matrices below:

$$P_1 = \begin{array}{c} \begin{array}{cccc} \text{Fresh.} & \text{Soph.} & \text{Junior} & \text{Senior} \end{array} \\ \left|\begin{array}{cccc} .9277 & 0 & 0 & 0 \\ .0005 & .8612 & 0 & 0 \\ 0 & .0313 & .9089 & 0 \\ 0 & 0 & .0047 & .7937 \end{array}\right| \end{array}$$

$$P_2 = \left|\begin{array}{cccc} .0964 & 0 & 0 & 0 \\ .6990 & .1001 & 0 & 0 \\ 0 & .7924 & .1393 & 0 \\ 0 & 0 & .7493 & .2917 \end{array}\right|$$

Entries in both of these matrices were estimated from Berkeley campus student performance and attrition data that has been collected and summarized by the Office of Institutional Research for the academic

year 1961. The second entry in each cell is calculated by the GPR method using the ratios

$$a_{11} = 0.0970$$
$$a_{12} = 0.7877$$
$$a_{23} = 0.8766$$
$$a_{34} = 0.9944$$

published by the State Department of Finance for the year 1960-1961. Initial enrollment and admission data for the forecast period are the same as those used in the GYA model. The third entry in each cell is the actual observed student enrollment. In Table I an italicized entry denotes which forecast is closer to the actual enrollment count.

We should point out that the interval between GPR forecasts was one year (Fall semester to Fall semester), while the forecast period for the GYA model was one semester; in Table I we only list the values appropriate to the beginning of the Fall semester. This fact in combination with the intrinsically larger forecast variances of the GPR model for $a_{j,j+1} > 1$ would seem to account for the discrepancies in the junior year.

In applying data to these models we have used the following convention: a <u>new</u> student is one coming to the U.C. Berkeley campus for the first time; in other words, he has not been registered before. A <u>continuing</u> student is one who has, with the exception of the summer sessions, a record of continuous registrations. For example, a student

TABLE I. Fall Semester Forecasts and Actual Enrollments* -- Berkeley Campus, 1962-1966

Class	1961	Model	1962	1963	1964	1965	1966
Freshman	3843	GYA	*3903*	*4013*	*3835*	2958	*3364*
		GPR	3902	4012	3833	*2963*	3360
		Actual	3972	4106	4186	3307	3633
Sophomore	3778	GYA	3746	*3834*	3805	*3408*	*3178*
		GPR	*3709*	3806	*3768*	3416	3016
		Actual	3649	3846	3468	3349	3126
Junior	4180	GYA	*4765*	4937	*4893*	*4475*	*4547*
		GPR	4728	*4819*	4779	4339	4400
		Actual	4762	4806	5429	5311	5624
Senior	3943	GYA	*4274*	*4767*	*4907*	5043	4882
		GPR	4341	4897	4994	*4879*	*4520*
		Actual	4210	4789	4585	4581	4364

* Actual Enrollments based on campus data from the Office of Institutional Research.

TABLE II. New Admissions* on Berkeley Campus, Spring 1962 - Fall 1966

	Spr '62	Fall '62	Spr '63	Fall '63	Spr '64	Fall '64	Spr '65	Fall '65	Spr '66	Fall '66
Freshmen	324	3,528	328	3,632	346	3,443	256	2,590	291	3,072
Sophomores	204	678	187	732	209	609	180	396	210	742
Juniors	303	1,416	324	1,568	173	1,443	452	1,035	476	1,406
Seniors	30	184	42	196	45	202	49	126	66	205

* A new student is defined as one who has never before registered at the Berkeley campus.

registered in the Spring semester of 1963 is a continuing student if he registers in Fall 1963. A returning student is a student who was once registered at the campus but has left for one or more semesters.

In calculating the diagonal entries of P_1 and P_2 we divided the number of continuing and returning students in one semester by enrollments of the previous semester. Clearly, from the definitions given above it would be better to relate returning students to enrollments of a semester two or more periods in the past.

BIBLIOGRAPHY

Bartholomew, D. J. (1967) Stochastic Models for Social Processess. J. Wiley.

Gani, J. (1963) Formulae for Projecting Enrolments and Degrees Awarded in Universities. J. R. Statist. Soc. A. 126, 400-409.

Pollard, J. H. (1967) Hierarchical Population Models with Poisson Recruitment. J. Appl. Prob. 4, 209-213.

Young, A. and Almond, G. (1961) Predicting Distributions of Staff. Comput. J. 3, No. 4, 246-250.

PUBLISHED REPORTS

68–3 Oliver, R. M., *Models for Predicting Gross Enrollments at the University of California.*

69–1 Marshall, K., and R. M. Oliver, *A Constant Work Model for Student Attendance and Enrollment.*

69–4 Breneman, D. W., *The Stability of Faculty Input Coefficients in Linear Workload Models of the University of California.*

69–10 Oliver, R. M., *An Equilibrium Model of Faculty Appointments, Promotions, and Quota Restrictions.*

P–1 Leimkuhler, F., and M. Cooper, *Analytical Planning for University Libraries.*

P–2 Leimkuhler, F., and M. Cooper, *Cost Accounting and Analysis for University Libraries.*

P–3 Sanderson, R. D., *The Expansion of University Facilities to Accommodate Increasing Enrollments.*

P–4 Bartholomew, D. J., *A Mathematical Analysis of Structural Control in a Graded Manpower System.*

P–5 Balderston, F. E., *Thinking About the Outputs of Higher Education.*

P–6 Weathersby, G. B., *Educational Planning and Decision Making: The Use of Decision and Control Analysis.*

P–7 Keller, J. E., *Higher Education Objectives: Measures of Performance and Effectiveness.*

P–8 Breneman, D. W., *An Economic Theory of Ph.D. Production.*

P–9 Winslow, F. D., *The Capital Costs of a University.*

P–10 Halpern, J., *Bounds for New Faculty Positions in a Budget Plan.*

P–11 Rowe, S., W. G. Wagner, and G. B. Weathersby, *A Control Theory Solution to Optimal Faculty Staffing.*

P–12 Weathersby, G. B., and M. C. Weinstein, *A Structural Comparison of Analytical Models.*

P–13 Pugliaresi, L. S., *Inquiries into a New Degree: The Candidate in Philosophy.*

P–14 Adams, R. F., and J. B. Michaelsen, *Assessing the Benefits of Collegiate Structure: The Case at Santa Cruz.*

P–15 Balderston, F. E., *The Repayment Period for Loan-Financed College Education.*

P–16 Breneman, D. W., *The Ph.D. Production Function: The Case at Berkeley.*

P–17 Breneman, D. W., *The Ph.D. Degree at Berkeley: Interviews, Placement, and Recommendations.*

P–18 Llubia, L., *An Analysis of the Schools of Business Administration at the University of California, Berkeley.*

P–19 Wing, P., *Costs of Medical Education.*

P–20 Kreplin, H. S., *Credit by Examination: A Review and Analysis of the Literature.*

P–21 Perl, L. J., *Graduation, Graduate School Attendance, and Investments in College Training.*

P–22 Wagner, W. G., and G. B. Weathersby, *Optimality in College Planning: A Control Theoretic Approach.*

P–23 Jewett, J. E., *College Admissions Planning: Use of a Student Segmentation Model.*

P–24 Breneman, D. W., (Editor), *Internal Pricing within the University—A Conference Report.*

P–25 Geoffrion, A. M., Dyer, J. S., and A. Feinberg, *Academic Departmental Management: An Application of an Interactive Multi-criterion Optimization Approach.*

P–26 Balderston, F. E., and R. Radner, *Academic Demand for New Ph.D.'s, 1970–90: Its Sensitivity to Alternative Policies.*

P–27 Morris, J., *Educational Training and Careers of Ph.D. Holders: An Exploratory Empirical Study.*

P–28 Wing, P., *Planning and Decision Making for Medical Education: An Analysis of Costs and Benefits.*

P–29 Balderston, F. E., *Varieties of Financial Crisis.*

P–30 Weathersby, G. B., *Structural Issues in the Supply and Demand for Scientific Manpower: Implications for National Manpower Policy.*

P–31 Balderston, F. E., and G. B. Weathersby, *PPBS in Higher Education Planning and Management from PPBS to Policy Analysis.*

P–32 Balderston, F. E., *Financing Postsecondary Education Statement to the Joint Committee on the Master Plan for Higher Education of the California Legislature, April 12, 1972.*

P–33 Balderston, F. E., *Cost Analysis in Higher Education.*

P–34 Smith, D. E., and W. G. Wagner, *SPACE: Space Planning and Cost Estimating Model for Higher Education.*

Single copies are available upon request; multiple copies available at cost. A price list may be obtained from the Ford Foundation office at: 2288 Fulton Street, Berkeley, California 94720.

E-524